

1. Descripció del Projecte

En els darrers anys hem assistit a una important evolució tecnològica, tant en la societat com en les organitzacions.

Actualment, pren gran rellevància en el si de les organitzacions, la gestió documental. La gestió documental la podem definir com: “el conjunt d'operacions i tècniques relatives a: la concepció, el desenvolupament, la implantació i l'avaluació dels sistemes administratius necessaris, des de la creació dels documents fins la seva destrucció o transferència a l'arxiu, per tal de garantir l'eficàcia i l'eficiència en la gestió administrativa”.

La incorporació d'un bon sistema de gestió documental redundarà en beneficis econòmics en les organitzacions, ja que facilita la identificació, la quantificació i la disposició dels documents; permetent economitzar en recursos materials i humans. Això ens porta a una millor organització i gestió de la informació, encaminada a optimitzar novament recursos humans i tecnològics.

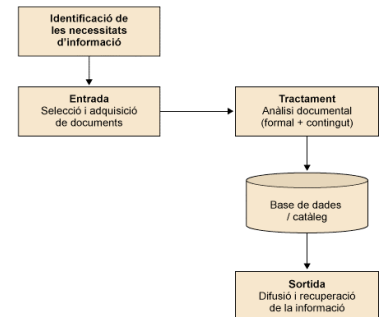


Diagrama 1, cicle de vida dels documents.

El Grup PH neix el desembre de 1990. És una empresa de base tecnològica amb àmplia experiència i amb vocació de servei; ofereix solucions informàtiques que permeten: optimitzar els processos empresarials, millorar la qualitat i l'accés a la informació, i reduir les despeses de gestió.

Tenim així que PH Systems disposa fins aquests moments d'un conjunt d'aplicacions informàtiques que conformen una solució de gestió de negoci integrada pensada per diversos sectors. Però aquestes aplicacions no disposen d'una eina que pugui fer una gestió de la documentació d'una organització.

2. Objectius

L'objectiu d'aquest projecte és millorar la gestió i/o la productivitat dels procediments informatitzats, que tenen com a element inicial en el seu flux de dades, els actius d'informació.

Això es pot subdividir en diferents objectius concrets:

- La utilització d'un Sistema d'Informació que permeti accedir i recuperar els documents d'arxiu d'una organització, sense haver d'anar a l'arxiu físic.
- Facilitar un procediment automatitzat, àgil i versàtil de creació de documents electrònics a partir dels documents d'arxiu.
- La integració dels actius d'informació amb la gestió de negoci que proporciona un ERP com el Lirba.

A part d'aquests objectius, també n'hi ha uns de secundaris:

- Dotar de semàntica els documents electrònics corresponents a documents administratius.
- Establir lligams entre aquesta semàntica i un flux de treball del Lirba.
- Permetre al Lirba treballar amb més d'un gestor de documents electrònics.

3. Records Management

El “**Records Management**” és una disciplina que estudia el tractament adequat en les organitzacions de certs actius d'informació denominats “**Records**”. Un Record és un actiu d'informació, digital o en paper, que reflecteix la història o memòria de l'organització respecte les seves activitats. Exemples d'aquests actius són: factures, comandes, documents comptables, etc.

Els actius d'informació es caracteritzen per:

- No haurien de ser alterables un cop creats.
- Existeix algun tipus d'exigència legal o corporativa perquè siguin conservats durant un període mínim de temps.
- Tenen valor per a l'organització.

4. Sistemes de Recuperació de la Informació

La Recuperació d'informació (RI a partir d'ara) és el procés de selecció d'informació d'un fons documental per mitjans automàtics o semiautomàtics, és a dir, amb l'ús parcial o intensiu d'ordinadors.

Tota la teoria moderna de la RI es fonamenta en les tres idees nuclears següents:

- La representació de la informació que contenen els diversos documents es realitza mitjançant l'assignació de diversos conjunts de termes d'indexació a cada document, i no tant per assignació dels documents a classes o subclasses d'un quadre de classificació.
- Les necessitats d'informació dels usuaris d'un sistema documental també es poden representar mitjançant conjunts de termes d'indexació.
- Els documents més rellevants amb relació a cada necessitat d'informació dels usuaris seran aquells que presentin un grau de semblança més gran respecte a la necessitat d'informació.

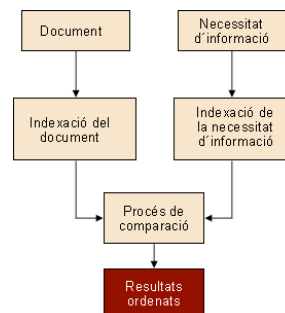


Diagrama 2, representació d'un Sistema R.I.

En un Sistema de RI no es poden comparar directament documents i necessitats d'informació. En realitat, el que es compara són representacions de cadascuna de les dues entitats esmentades.

5. Utilització de les TIC

Les tecnologies de la informació tenen, d'una banda, una faceta d'automatització, és a dir, resulten molt eficients quan ens reemplacen en tasques rutinàries, en operacions que es poden reduir a operacions numèriques o lògiques. En aquest sentit, les aplicacions d'automatització han tingut un impacte important en la indústria, resultant en un augment de la productivitat. D'altra banda, les tecnologies de la informació tenen també una faceta d'augmentació, és a dir, poden multiplicar les nostres capacitats intel·lectuals, ens poden ajudar a pensar, a escriure, a decidir i a actuar millor.

Un cas exemplificatiu seria: el fet que l'autor utilitzi un ordinador per a escriure aquest text no representa una aplicació augmentadora de les seves capacitats intel·lectuals; ho seria, per exemple, si la màquina fos capaç d'ordenar les idees disperses introduïdes per l'autor i les convertís directament en un text no solament llegible, sinó intel·ligent i tot.

6. Requeriments del projecte

Una mancança que té el Lirba és que no fa cap gestió dels documents administratius. Com a actius d'informació que són, concretament informació formal i estructurada; podem amb les TIC realitzar una automatització i fins i tot una augmentació del tractament de la informació.

Per tant, de manera global s'haurà de procurar:

- Que els components es puguin configurar segons altres SI preexistents.
- Que es disposi de més d'una alternativa d'implantació i de funcionament de les principals funcionalitats.
- Que els components del sistema puguin ser distribuïts i això es pugui determinar en la implantació o en futures actualitzacions.
- Que hi hagi un baix acoblament entre els components per facilitar la creació de paquets i la seva implantació.
- Que les funcionalitats proveïdes per terceres aplicacions es puguin reemplaçar sense afectar el sistema, ni la seva funcionalitat.
- Que hi hagi un alt grau de reutilització.

També s'ha de valorar la normativa de la Agència Tributària que permet l'eliminació de factures físiques després de realitzar un procés de digitalització certificada.

Així doncs tindrem tres aspectes fonamentals a tractar:

- Els documents administratius o actius d'informació.
- L'adquisició d'aquests documents i la gestió del sistema.
- La integració del Lirba amb un Record Management.

7. Anàlisi del projecte

El projecte està compost per tres components principals: l'Arxiu Òptic, l'aplicació d'administració i adquisició de documents i el Lirba. Aquests components principals poden interactuar també amb altres components externs.

Els diferents components que hi ha són:

- **Arxiu Òptic** :- Aquest component s'encarrega de la gestió dels documents administratius, entre altres tasques té la funció d'Indexació i classificació dels documents, així com la de generar la seva metainformació.
- **Administració i Adquisició de documents** :- Aquesta aplicació proporciona una interfície per administrar l'Arxiu Òptic, així com un mètode d'adquisició de documents en format físic
- **Lirba** :- Aquest paquet informàtic s'encarrega de la gestió d'una organització, a més ara integrarà dins dels seus fluxos de treball els documents administratius.

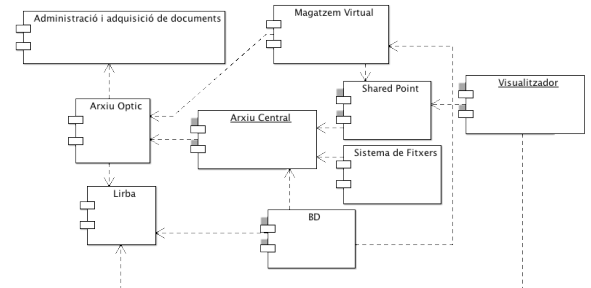


Diagrama 3, model general amb els tres components principals.

8. Disseny del projecte

Per tal d'assegurar la reusabilitat en el disseny dels elements del sistema, així com evitar la reiteració a la recerca de solucions ja conegudes i emprades anteriorment s'han utilitzat diferents patrons, tant *GRASP* com *GoF*.

L'«*Abstrac Factory*» permet treballar amb diferents tipologies de famílies d'objectes, s'ha emprat profusament per poder desacoplar, principalment en els components principals de l'Arxiu Òptic. Així doncs, tots els components, s'han estructurat sobre col·leccions abstractes i d'aquesta manera, el sistema interactua amb un únic tipus d'element, independentment de la funcionalitat que tingui o la persistència que se li vulgui donar.

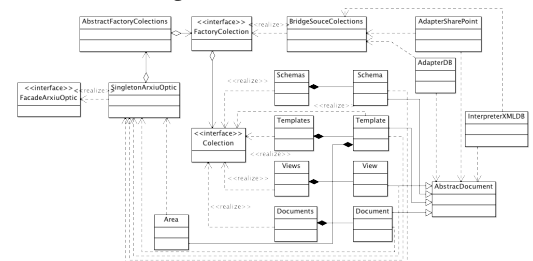


Diagrama 4, model de classes del component remot.

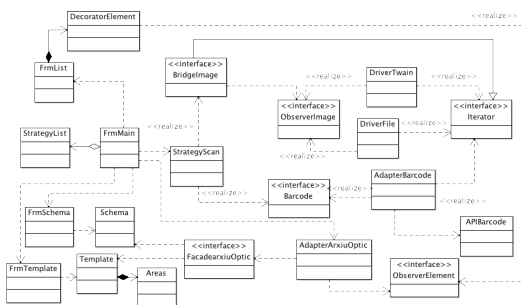


Diagrama 5, model de classes aplicació admin. i adq. Documents.

el sistema de fitxers.

L'«*Iterator*» permet realitzar recorreguts sobre objectes de forma independent de la seva implementació, s'ha emprat profusament, ja que tot el sistema treballa sobre col·leccions. També s'ha emprat en la creació del Driver Twain.

L'«*Observer*» defineix una relació d'un a molts, permetent la notificació de canvis d'estat. S'ha emprat per a poder treballar asíncronament, o bé per treballar síncronament a partir de sistemes asíncrons. Juntament amb l'«*Strategy*», l'«*Iterator*», l'«*Adapter*» i el «*Bridge*» conformen la funcionalitat d'adquisició d'imatges.

El patró «*Strategy*» permet la utilització de diferents mètodes per resoldre un problema, s'ha emprat, per resoldre la creació de documents, permetent l'adquisició d'imatges des de diferents tipus de fons origen.

L'«*Adapter*» adapta un objecte ja existent a una funcionalitat no prevista, s'ha emprat entre altres, per adaptar la funcionalitat de productes de terceres companyies a funcionalitats del sistema, com els serveis Web del SharePoint a la funcionalitat de l'Arxiu Òptic o bé el reconeixement de codis de barres.

El «*Bridge*» desacopla una abstracció de la seva implementació, s'ha emprat entre altres per l'adquisició de documents, permetent realitzar aquesta a partir de diferents fons externes, com un escàner o bé

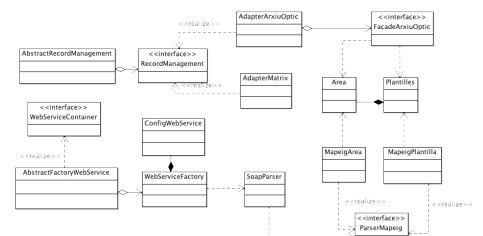


Diagrama 6, model de classes integració amb el Lirba.

9. Cas: Entrada d'assentaments comptables al Lirba

Lirba Financials està pensat per a la petita i mitjana empresa, agilitzant i ajudant a resoldre la casuística comptable d'una manera ràpida i intuïtiva, sota un entorn senzill i amigable, ahora que robust.

Aquesta comptabilitat està pensada per introduir documents enlloc d'assentaments, per tant el primer que cal fer, és definir els documents comptables.

Imatge 1, definició de documents comptables al Lirba.

Imatge 2, definició esquemes Arxiu Òptic.

Ahora també s'han de definir els documents a l'Arxiu Òptic. S'ha de definir un esquema (metadades), i una plantilla. Aquesta relaciona regions d'un document amb URIS d'un esquema. Això permetrà posteriorment, la generació d'un fitxer XML, validat per l'esquema introduït, i que conté les dades del document escanejat gràcies a un OCR.

Abans, d'introduir els assentaments s'ha d'establir el lligam entre el Lirba i l'Arxiu Òptic. Hi ha dos nivells, la definició genèrica (vàlida per altres aplicacions) i la definició dels vincles entre els flux de treball i els Uris dels esquemes (pròpia de l'Arxiu Òptic).

Imatge 3, definició Plantilles Arxiu Òptic.

Imatge 4, definició flux treball Lirba.

Imatge 5, definició variables flux treball Lirba.

Imatge 6, vincle flux esquema.

Imatge 7, vincle entre variables flux i Uris esquema.

Un cop fet el lligam, això ens permetrà, al introduir un assentament comptable d'un tipus de document, escollir entre els documents de l'Arxiu Òptic pendents de processar, per tal d'introduir de forma automàtica la informació generada (en format XML) en les corresponents caselles del formulari.

Imatge 8, entrada assentaments comptables.

Imatge 9, selecció documents Arxiu Òptic pendents de processar.

Imatge 10, dades introduïdes de forma automàtica.

D'aquesta manera, s'estableix una relació, entre una instància d'un procés (en aquest cas un assentament comptable) amb un document de l'Arxiu Òptic. Sempre que es recuperi l'assentament es podrà visualitzar el document.

Imatge 11, visualització d'un document des d'un apunt comptable.

Amb aquest exemple, veiem com no només s'ha realitzat una automatització (gestió dels actius de informació), sinó que també hi ha hagut una augmentació (generació automàtica d'apunts comptables).

10. Conclusions

Aquest projecte pretenia transformar una mancança (la falta de la gestió dels actius d'informació) detectada en el conjunt d'aplicacions informàtiques de PH Systems; en una fortaleesa (la millora de la productivitat).

Per aconseguir això, s'havien plantejat diferents objectius concrets. A més, conforme s'han establert els requeriments i s'ha realitzat l'anàlisi, s'han detectat noves oportunitats que s'han resolt satisfactòriament; aconseguint amb això la automatització de la gestió dels actius de informació. Així doncs:

- El conjunt d'eines realitzades permeten accedir i recuperar documents d'arxiu corrent (Record Management). A més, s'han seguit les directives que permeten (si es vol) l'eliminació dels documents físics.
- De les eines realitzades, destaca l'aplicació d'administració, que funciona com a Arxiu Òptic, permetent de forma automatitzada, àgil i versàtil la creació de documents electrònics a partir dels documents d'arxiu corrent.
- També s'ha realitzat la integració dels actius d'informació amb la gestió de negoci que proporciona un ERP com el Lirba; on s'ha tingut especial cura en proporcionar la funcionalitat sense acoblar-la a les eines realitzades, permetent la utilització de productes de terceres companyies (com és el cas de Matrix o DocuWare)

La consecució dels objectius secundaris (concretament la de dos d'ells) ha estat cabdal pel projecte, ja que l'ha dotat d'una faceta augmentadora:

- Amb la incorporació dels Esquemes XML i les plantilles, s'ha pogut resoldre la problemàtica de dotar de semàntica els documents electrònics corresponents a documents administratius.
- Aprofitant el coneixement del Nucli del Lirba i la poderosa eina que ens proporciona el repositori d'aquest, s'ha pogut establir lligams entre semàntica dels documents d'arxiu corrent i els fluxos de treball del Lirba.

Un cop vistes les fortaleeses de les eines que s'han generat (gràcies a les oportunitats detectades), hi ha una sèrie de febleses que s'haurien de millorar:

- Existeix una forta limitació en quan a la generació de les metadades, aquesta limitació està causada per:
 - La utilització de la llibreria Modi per la realització de la tasca de OCR.
 - La utilització de regions basades en píxels per la creació de les plantilles.
- En aquest projecte s'ha deixat totalment de banda la gestió documental per centrar-nos en els documents d'arxiu corrent, i per tant no s'han tingut en compte documents del tipus: patents, informació tècnica, etc... Que sí són necessaris en determinats departaments d'una organització.
- S'ha fet un gran esforç per a realitzar un sistema d'indexació flexible, i emprar un llenguatge basat en les metadades per tal de formular les necessitats d'informació; però s'ha deixat tota la feina del llenguatge d'interrogació a l'ús de terceres aplicacions, com seria el cas d'emprar una base de dades nativa en XML per a contenir les metadades dels documents electrònics generats.

Aquestes mancances o febleses, s'han intentat resoldre a partir de l'acció de l'usuari o bé amb el coneixement de les regles de negoci, així per exemple en el cas de la comptabilitat:

- No es requereix una complexa formulació de les necessitats d'informació, ja que tots els documents que pertanyen a un tipus concret, si no s'han processat prèviament s'han de processar obligatòriament (s'han de comptabilitzar totes les factures).
- Si l'OCR no funciona correctament sempre l'usuari pot modificar la dada incorrecta i aquest canvi modificarà les metadades del document electrònic (afinant l'OCR).
- No és necessari una complexa detecció de regions o àrees, amb la capçalera i el peu d'una factura n'hauria d'haver-hi prou.

En qualsevol cas i tenint en compte, la modularitat i extensibilitat del sistema desenvolupat, s'han de considerar com a futures aportacions:

- La utilització d'eines de tercers o pròpies que permetin una detecció de plantilles i de les regions o àrees d'aquestes de forma automàtica (i no basada en píxels), de tal manera que el sistema aprengui a partir de la inclusió de nous documents.
- La millor integració amb gestors de continguts o gestions documentals pel tractament de documents extensos.